Measures of Quality of Rulesets Extracted from Data

Martin Holeňa

Institute of Computer Science, Academy of Sciences of the Czech Republic,
Pod vodárenskou věží 2, 18207 Praha 8, Czech Republic, martin@cs.cas.cz, web:cs.cas.cz/~martin

Abstract. *The paper deals with quality measures of whole sets of rules extracted from data, as a counterpart to more commonly used measures of individual rules. This research has been motivated by increasingly frequent extraction of non-classification rules, such as association rules and rules of observational logic, in real-world data mining tasks. The paer sketches the typology of rules extraction methods and of their rulesets, and recalls that quality measures for whole sets of rules have been so far used only in the case of classification rulesets. It then proposes three possible ways how such measures can be extended to general rulesets. The paper also recalls the possibility to measure the dependence of classification ruleset on parameters of the classification method by means of ROC curves, and proposes a generalization of ROC curves to general rulesets. Finally, a brief illustration on rulesets extracted by means of the method GUHA is given.*

1 Introduction

Logical formulas of specific kinds, usually called *rules*, are a traditional way of formally representing knowledge. Therefore, it is not surprising that they are also the most frequent representation of the knowledge discovered in data mining. Existing methods for rules extraction are based on a broad variety of paradigms and theoretical principles. However, methods relying on different underlying assumptions can lead to the extraction of different or even contradictory rulesets from the same data. Moreover, the set of rules extracted with a particular method can substantially depend on some tunable parameter or parameters of the method, such as significance level, thresholds, size parameters, trade-off coefficients etc. For that reason, it is desirable to have measures of various qualitative aspects of the extracted rulesets. So far, such measures are available only for sets of classification rules, and their dependence on tunable parameters can be described only for classification into two classes [10, 15]. As far as more general kinds of rules are concerned, measures of quality have been proposed only for individual rules [6, 11, 24, 26, 29], or for contrast sets of rules, which finally can be replaced with a single rule [2, 16]; if a whole ruleset is taken into consideration, then only as a context for measuring the quality of an individual rule [27, 28].

The research reporeted in this paper has been motivated by increasingly frequent extraction of non-classification rules in real-world data mining tasks. The paper discusses three possible ways of extending existing ruleset quality measures from classification to general rulesets. The proposed extensions are introduced in Section 4, after the basic typology of rules extraction methods and examples of measures for classification rulesets are recalled in the following two sections, and before a generalization of ROC curves is proposed in Section 5. The paper concludes with a brief illustration on rulesets extracted with the method GUHA.

2 Typology of rules extraction methods

The most natural base for differentiating between existing rules extraction methods is the *syntax and semantics of the extracted rules*. Syntactical differences between them are, however, not very deep since principally, any rule r has one of the forms $S_r \sim S'_r$, or $A_r \to C_r$, where S_r, S'_r, A_r and C_r are formulas of the considered logic, and $\sim$, $\to$ are symbols of the language of that logic. The difference between both forms concerns semantic properties of the symbols $\sim$ and $\to$: $S_r \sim S'_r$ is symmetric with respect to S_r, S'_r in the sense that its validity always coincides with that of $S_r \sim S'_r$ whereas $A_r \to C_r$ is not symmetric with respect to A_r, C_r in that sense. In the case of a propositional logic, $\sim$ and $\to$ are the connectives equivalence and implication, respectively, whereas in the case of a predicate logic, they are generalized quantifiers. To distinguish the formulas involved in the asymmetric case, A_r is called *antecedent* and C_r *consequent* of r.

The more important is the semantic of the rules (cf. [6]), especially the difference between *rules of the Boolean logic* and *rules of a fuzzy logic*. Due to the semantics of Boolean and fuzzy formulas, the former are valid for crisp sets of objects, whereas the validity of the latter is a fuzzy set on the universe of all considered objects. Boolean rulesets are extracted more frequently, especially some specific types of them, such as *classification rulesets* [11, 15]. Those are sets of implications such that $(A_r)_{r\in\mathcal{R}}$ and $\{C_r\}_{r\in\mathcal{R}}$ partition the set $\mathcal{O}$ of considered objects, where $\mathcal{R}$ is the considered ruleset, and $\{C_r\}_{r\in\mathcal{R}}$ stands for the set of distinct formulas in $(C_r)_{r\in\mathcal{R}}$. Abandoning the requirement that $(A_r)_{r\in\mathcal{R}}$ partitions $\mathcal{O}$ (at least in the sense of a crisp partitioning) allows to generalize those rulesets also to fuzzy antecedents. For Boolean antecedents, however, this requirement entails a natural definition of the validity of a whole classification ruleset $\mathcal{R}$ for an object x. Assuming that all information about x conveyed by $\mathcal{R}$ is conveyed by the single rule r covering x (i.e., with A_r valid for x), the validity of $\mathcal{R}$ for x can be defined to coincide with the validity of $A_r \to C_r$ for that r, which in turn equals the validity of C_r for x.

As far as the Boolean predicate logic is concerned, generalized quantifiers both for symmetric and for asymmetric rules were studied in the 1970s within the framework of the *observational logic* [13], which is a Boolean predicate logic with generalized quantifiers. For a set of data about n objects, the truth evaluation of the Boolean predicate φ on those objects is a vector $\|\varphi\| \in \{0,1\}^n$, whereas the truth evaluation of a sentence $(Qx)(\varphi_1(x), \dots, \varphi_m(x))$ consisting of m Boolean predicates $\varphi_1, \dots, \varphi_m$ and an m-ary generalized quantifier Q is the function value

$$\|(Qx)(\varphi_1(x), \dots, \varphi_m(x))\| = \mathrm{Tf}_Q(\|\varphi_1\|, \dots, \|\varphi_m\|), \quad (1)$$

of a $\{0,1\}$-valued function Tf_Q on the set of m-column binary matrices, which is called *truth function* of the quantifier Q. Observational logic underlies one of the earliest methods for the extraction of general rules from data, called General Unary Hypotheses Automaton (GUHA). In GUHA, the truth function Tf_Q of a generalized quantifier Q is always a function of the 4-fold table

$$\begin{array}{cc|cc} & & S'_r & \neg S'_r \\ \cline{3-4} & & C_r & \neg C_r \\ \hline S_r & A_r & a & b \\ \neg S_r & \neg A_r & c & d \end{array}. \quad (2)$$

Hence, Tf_Q is a $\{0,1\}$-valued function on quadruples of nonnegative integers. For symmetric rules, GUHA uses quantifiers fulfilling

$$\begin{aligned} &a' \geq a \;\&\; b' \leq b \;\&\; c' \leq c \;\&\; d' \geq d \;\& \\ &\;\&\; \mathrm{Tf}_Q(a,b,c,d) = 1 \to \mathrm{Tf}_Q(a',b',c',d') = 1. \end{aligned} \quad (3)$$

They are called *associational quantifiers*. For asymmetric rules, it uses quantifiers fulfilling the stronger condition

$$\begin{aligned} &a' \geq a \;\&\; b' \leq b \;\& \\ &\;\&\; \mathrm{Tf}_Q(a,b,c,d) = 1 \to \mathrm{Tf}_Q(a',b',c',d') = 1. \end{aligned} \quad (4)$$

which are called *implicational quantifiers*. This condition covers also the frequently encountered *association rules* [1, 6, 40] (since methods for the extraction of association rules have been developed outside the framework of observational logic, the terminology is a bit confusing here: although associational rules are asymmetric, their name evokes the quantifier for the symmetric ones).

Orthogonally to the typology according to the semantics of the extracted rules, all extraction methods can be divided into two large groups:

– Methods that extract logical rules from data *directly*, without any intermediate formal representation of the discovered knowledge. Such methods have always formed the mainstream of the extraction of Boolean rules: from the observational logic methods [13] and the method *AQ* [30, 31] in the late 1970s, through the extraction of association rules [1, 40] and the method *CN2* [4], relying on a paradigm similar to that of AQ, to recent methods based on *inductive logic programming* [5, 33] and *genetic algorithms* [9]. They include also important methods for fuzzy rules, in particular *ANFIS* [22, 23] and *NEFCLASS* [34, 35], *fuzzy generalizations* of *observational logic* [18, 19] and a recent method based on *fuzzy transform* [36].
– Methods that employ some *intermediate representation* of the extracted knowledge, useful by itself. This group includes two important kinds of methods: *classification trees* [3, 37] and methods based on *artificial neural networks (ANN)*. The latter are used both for Boolean and for fuzzy rules [7, 21, 39] (cf. also the survey papers [32, 38]).

3 Existing measures for classification rulesets

A survey of measures of quality for classification rulesets (with possibly fuzzy antecedents) has been given in the monograph [15]. All measures have been divided there into four groups: inaccuracy, imprecision, inseparability and resemblance. Space limitation allows to recall here only the main representatives of the more important groups:

Inaccuracy measures the discrepancy between the true class of the considered objects and the class predicted by the ruleset. Its most frequently encountered representative is the *quadratic score* (also called Brier score):

$$\mathrm{Inacc} = \frac{1}{|\mathcal{O}|} \sum_{x \in \mathcal{O}} \sum_{C \in \{C_r\}_{r \in \mathcal{R}}} \left(\delta_C(x) - \hat{\delta}_C(x)\right)^2, \quad (5)$$

where $|\,|$ denotes cardinality, $\mathcal{O}$ is the considered set of objects, $\delta_C(x) \in \{0,1\}$ is the validity of the proposition C for $x \in \mathcal{O}$, and $\hat{\delta}_C(x)$ is the agreement between C and the class predicted for x by $\mathcal{R}$. In the general case of a fuzzy logic, $\hat{\delta}_C(x) = \max_{C_r=C} \|A_r\|_x$, with $\|A_r\|_x \in \langle 0,1\rangle$ denoting the truth grade of A_r for x.

Imprecision measures the discrepancy between the probability distribution of the classes, conditioned on the values of attributes occurring in antecedents, and the class predicted by the ruleset. Its most common representative is

$$\text{Impr} = \\ = \frac{1}{|\mathcal{O}|} \sum_{x\in\mathcal{O}} \sum_{C\in\{C_r\}_{r\in\mathcal{R}}} \left(\delta_C(x) - \hat{\delta}_C(x)\right)\left(1-\hat{\delta}_C(x)\right)^2. \tag{6}$$

As was already mentioned in the introduction, the extracted ruleset can substantially depend on tunable parameters of the employed method. This was so far systematically studied only for dichotomous classification with $\mathcal{R} = \{A \to C, \neg A \to \neg C\}$. In that case, putting $A_r = A, C_r = C$ allows the information about the validity of A and C for $\mathcal{O}$ to be again summarized by means of the 4-fold table (2), which also depends on the parameter values. The influence of the parameter values on the result of dichotomous classification is usually investigated by means of the measures *sensitivity* $= \frac{a}{a+c}$ and *specificity* $= \frac{d}{b+d}$ [15]. Connecting points (1-specificity,sensitivity) $= (\frac{b}{b+d}, \frac{a}{a+c})$ for the considered parameter values forms a curve with graph in the unit square, called *receiver operating characteristic* (ROC), due to the area where such curves have first been in routine use. In machine learning, a modified version of those curves has been proposed, in which the points connected for considered parameter values are (b, a) [10]. The graph of such a curve then lies in the rectangle with vertices $(0,0)$ and $(b+d, a+c)$, and is called *coverage graph*.

The graphs of ROC curves and coverage graphs can provide information about the influence of parameter values not only on the sensitivity and specificity, but also on other measures. It is sufficient to complement the graph with isolines of the measure and to investigate their intersections with the original curve [10].

4 Three extensions to more general kinds of rules

In the particular case of classification rulesets with Boolean antecedents, some algebra allows to substantially simplify (5)–(6):

$$\begin{aligned} \text{Inacc} &= \frac{2|\mathcal{O}^-|}{|\mathcal{O}|} = 1 - \frac{|\mathcal{O}^+| - |\mathcal{O}^-|}{|\mathcal{O}|}, \\ \text{Impr} &= \frac{|\mathcal{O}^-|}{|\mathcal{O}|} = 1 - \frac{|\mathcal{O}^+|}{|\mathcal{O}|}, \end{aligned} \tag{7}$$

where

$$\begin{aligned} \mathcal{O}^+ &= \{x \in \mathcal{O} : \mathcal{R} \text{ is valid for } x\}, \\ \mathcal{O}^- &= \{x \in \mathcal{O} : \mathcal{R} \text{ is not valid for } x\}. \end{aligned} \tag{8}$$

This not only shows that, in the case of Boolean antecedents, the quadratic score is sufficient to describe also the imprecision, but also suggests an approach how to extend those measures to general rulesets: to use (7)–(8) as the definition of measures (5)–(6). More generally, any measure of quality of classification rulesets with Boolean antecedents (e.g., any measure surveyed in [15]) that can be reformulated by means of $\mathcal{O}^+$ and $\mathcal{O}^-$, can be extended in such a way that the reformulation is used as the definition of that measure for general rulesets.

For sets of asymmetric rules, also the notion of covering an object by a rule, which was recalled in Section 2, can be generalized. Notice, however, that for fuzzy antecedents, the validity of A_r, $r \in \mathcal{R}$ is a fuzzy set on $\mathcal{O}$. Consequently, the set $\mathcal{O}_\mathcal{R}$ of objects covered by $\mathcal{R}$ is a fuzzy set on $\mathcal{O}$ with the membership function

$$\mu_\mathcal{R}(x) = \|(\exists r \in \mathcal{R})\ A_r\|_x = \max_{r\in\mathcal{R}} \|A_r\|_x. \tag{9}$$

Observe that according to (9), $\mathcal{O}_\mathcal{R} = \mathcal{O}$ for classification rulesets with Boolean antecedents. Therefore, various generalizations of classification measures to general rulesets of asymmetric rules are possible: wherever $\mathcal{O}$ occurs in the definition of a measure for classification rulesets, either $\mathcal{O}$ or $\mathcal{O}_\mathcal{R}$ can occur in its general definition, provided $\mathcal{O}_\mathcal{R} \neq \emptyset$. To allow unified treatment of symmetric and asymmetric rules, the concept of covering an object by a rule will be extended also to symmetric rules, in such a way that an object x is covered by $S_r \sim S'_r$ if either S_r or S'_r is valid for x. Hence, a counterpart of (9) for a set $\mathcal{R}$ is a fuzzy set with the membership function

$$\begin{aligned} \mu_\mathcal{R}(x) &= \|(\exists r \in \mathcal{R})(S_r \vee S'_r)\|_x = \\ &= \max_{r\in\mathcal{R}} \max(\|S_r\|_x, \|S'_r\|_x). \end{aligned} \tag{10}$$

According to (8), the proposed way of extending measures of quality from classification rulesets with Boolean antecedents to general rulesets requires to generalize the concept of validity of a general ruleset for an object. However, there are multiple possibilities for such a generalization. Indeed, at least any of the following points of view is possible:

Boolean validity of the ruleset based on simultaneous validity of all covering rules. According to this point of view, the validity of a ruleset $\mathcal{R}$ for a covered object x is a Boolean property expressing the simultaneous validity of all rules that cover x.

Consequently, the sets $\mathcal{O}^+$ and $\mathcal{O}^-$ defined in (8) are crisp sets

$$\mathcal{O}^+ = \{x \in \mathcal{O} : \mu_{\mathcal{R}}(x) > 0 \;\&\; (\forall r \in \mathcal{R})\; \|r \text{ covers } x \;\&\; r \text{ is valid for } x\| = \|r \text{ covers } x\|\}, \tag{11}$$

$$\mathcal{O}^- = \{x \in \mathcal{O} : \mu_{\mathcal{R}}(x) > 0 \;\&\; (\exists r \in \mathcal{R})\; \|r \text{ covers } x \;\&\; r \text{ is valid for } x\| < \|r \text{ covers } x\|\}, \tag{12}$$

where

$$\|r \text{ covers } x\| = \begin{cases} \|(S_r \vee S'_r)\|_x & \text{for symmetric rules ,} \\ \|A_r\|_x & \text{for asymmetric rules ,} \end{cases} \tag{13}$$

and similarly

$$\|r \text{ covers } x \;\&\; r \text{ is valid for } x\| = \\ = \begin{cases} \|(S_r \vee S'_r)\&r\|_x & \text{for symmetric rules ,} \\ \|A_r \& r\|_x & \text{for asymmetric rules .} \end{cases} \tag{14}$$

The following consequences of this point of view are worth noticing:

(i) It is immaterial how the truth grade $\|r\|_x$ of a rule r being valid for an object x is evaluated (thus also how $\|\neg r\|_x$ is evaluated).

(ii) If $\mu_{\mathcal{R}}(x) = 0$, then $x \notin \mathcal{O}^+ \cup \mathcal{O}^-$.

(iii) For classification rulesets with Boolean antecedents, the validity of $\mathcal{R}$ according to this point of view coincides with the definition in Section 2 because in that case, there is exactly one rule that covers x.

Boolean validity of the ruleset based on the validity of the majority of covering rules. According to this point of view, the validity of a ruleset $\mathcal{R}$ for a covered object x is a Boolean property expressing the validity of most of the rules that cover x. Consequently, the sets $\mathcal{O}^+$ and $\mathcal{O}^-$ in (8) are crisp sets

$$\mathcal{O}^+ = \{x \in \mathcal{O} : \mu_{\mathcal{R}}(x) > 0 \;\&\; \sum_{r \in \mathcal{R}} \|r \text{ covers } x \;\&\; r \text{ is valid for } x\| > \sum_{r \in \mathcal{R}} \|r \text{ covers } x \;\&\; \neg r \text{ is valid for } x\|\}, \tag{15}$$

$$\mathcal{O}^- = \{x \in \mathcal{O} : \mu_{\mathcal{R}}(x) > 0 \;\&\; \sum_{r \in \mathcal{R}} \|r \text{ covers } x \;\&\; r \text{ is valid for } x\| \leq | \sum_{r \in \mathcal{R}} \|r \text{ covers } x \;\&\; \neg r \text{ is valid for } x\|\}, \tag{16}$$

where the truth grade $\|r$ covers $\;\&\; \neg r$ is valid for $x\|$ is again evaluated according to (14), replacing r with $\neg r$. Observe that also this point of view has the above consequences (i)–(iii), the last one again due to the fact that there is exactly one rule covering x.

Fuzzy validity of the ruleset based on the relative validity of covering rules. In this case, the validity of a ruleset $\mathcal{R}$ for a covered object x is a fuzzy property expressing the ratio of the validity of rules from $\mathcal{R}$ for x to the covering of x with those rules. Consequently, the sets $\mathcal{O}^+$ and $\mathcal{O}^-$ are fuzzy sets on $\mathcal{O}$ with memberships μ_+ and μ_-, respectively, such that if $\mu_{\mathcal{R}}(x) > 0$,

$$\mu_+(x) = \frac{\sum_{r \in \mathcal{R}} \|r \text{ covers } x \;\&\; r \text{ is valid for } x\|}{\sum_{r \in \mathcal{R}} \|r \text{ covers } x\||} \tag{17}$$

$$\mu_-(x) = \frac{\sum_{r \in \mathcal{R}} \|r \text{ covers } x \;\&\; \neg r \text{ is valid for } x\|}{\sum_{r \in \mathcal{R}} \|r \text{ covers } x\||} \tag{18}$$

where the involved truth grades are again evaluated according to (13) and (14). Moreover, (17)–(18) will be complemented with the definition $\mu_+(x) = \mu_-(x) = 0$ if $\mu_{\mathcal{R}}(x) = 0$, to get again the validity of (ii) above, whereas (i) and (iii) are consequences also of this point of view. Further, the fact that $\mathcal{O}^+$ and $\mathcal{O}^-$ are now fuzzy sets implies that whenever $|\mathcal{O}^+|$ or $|\mathcal{O}^-|$ occur in the definitions of quality measures for Boolean classification rulesets, fuzzy cardinalities have to be used in their generalizations to general rulesets according to this point of view. Hence,

$$|\mathcal{O}^+| = \sum_{x \in \mathcal{O}} \mu_+(x),\; |\mathcal{O}^-| = \sum_{x \in \mathcal{O}} \mu_-(x). \tag{19}$$

For example, the measure

$$\text{Inacc} = 1 - \frac{\sum_{x \in \mathcal{O}} (\mu_+(x) - \mu_-(x))}{|\mathcal{O}|} \tag{20}$$

is a generalization of (5), whereas the measures

$$\text{Impr}_1 = 1 - \frac{\sum_{x \in \mathcal{O}} \mu_+(x)}{|\mathcal{O}|}, \tag{21}$$

$$\text{Impr}_2 = 1 - \frac{\sum_{x \in \mathcal{O}} \mu_+(x)}{|\mathcal{O}_{\mathcal{R}}|} = 1 - \frac{\sum_{x \in \mathcal{O}} \mu_+(x)}{\sum_{x \in \mathcal{O}} \mu_{\mathcal{R}}(x)} \tag{22}$$

are generalizations of (6).

5 Extensions of ROC curves to more general kinds of rules

Observe that in the case of Boolean classification with $\mathcal{R} = \{A \to C, \neg A \to \neg C\}$, the information about the validity of $\mathcal{R}$ for objects $x \in \mathcal{O}$ can be also viewed as information about the validity of a ruleset $\mathcal{R}' = \{A \rightarrow C\}$. However, $\mathcal{R}'$ is not any more a classification ruleset, but only a general one, which can be described only by means of the above introduced sets $\mathcal{O}_\mathcal{R}$, $\mathcal{O}^+$, $\mathcal{O}^-$. In particular, $|\mathcal{O}^+| = a$ and $|\mathcal{O}^-| = b$, which suggests the possibility to generalize coverage graphs introduced in Section 3 to general rulesets by means of a curve connecting points $(|\mathcal{O}^-|, |\mathcal{O}^+|)$ for each of the values of the considered parameters. For a generalization of ROC curves to general rulesets, those points have to be scaled to the unit square. Since the resulting curve will be used to investigate the dependence on parameter values, the scaling factor itself must be independent of those values. The only available factor fulfilling this condition is the number of objects, $|\mathcal{O}|$ (the other available factors, $|\mathcal{O}_\mathcal{R}|$, $|\mathcal{O}^+|$ and $|\mathcal{O}^-|$ depend on the evaluations $\|S_r\|$ and $\|S'_r\|$, or $\|A_r\|$ and $\|C_r\|$, which in turn depend on the parameter values). Consequently, the proposed generalization of ROC curves will connect points $(\frac{|\mathcal{O}^-|}{|\mathcal{O}|}, \frac{|\mathcal{O}^+|}{|\mathcal{O}|})$.

For practical construction of the proposed generalization of ROC curves, the following proposition, proven in [17], can be quite useful:

Proposition 1. *Let the covering of individual objects with individual rules be a Boolean property (i.e., the set of rules covering a particular object x be a crisp subset of $\mathcal{R}$). Then irrespectively of which of the above points of view of ruleset validity is adopted, there always exists a constant $c \in \langle 0, 1\rangle$ and an increasing bijection $g : \langle 0, c\rangle \rightarrow \langle 0, 1\rangle$ such that*

$$|\mathcal{O}^+| + |\mathcal{O}^-| \leq \max(1, \max_{x \in \langle 0,c\rangle} x + g^{-1}(1 - g(x)))|\mathcal{O}|. \quad (23)$$

Moreover, in the particular cases of Boolean logic and of all three fundamental fuzzy logics (Łukasiewicz, Gödel, product), (23) holds with $c = 1$ and g equal to identity,

$$|\mathcal{O}^+| + |\mathcal{O}^-| \leq |\mathcal{O}|. \quad (24)$$

Thus in those cases, the points $(\frac{|\mathcal{O}^-|}{|\mathcal{O}|}, \frac{|\mathcal{O}^+|}{|\mathcal{O}|})$, forming the generalization of ROC curves, lie below the diagonal $(\langle 0, 1\rangle, \langle 1, 0\rangle)$.

The proposition is illustrated in Figure 1, together with isolines of the three example measures introduced in (20)–(22). Observe that the isolines of Impr_2 depend on the relationship between the three cardinalities $|\mathcal{O}^+| = \sum_{x\in\mathcal{O}} \mu_+(x)$, $|\mathcal{O}^-| = \sum_{x\in\mathcal{O}} \mu_-(x)$ and $|\mathcal{O}_\mathcal{R}| = \sum_{x\in\mathcal{O}} \mu_\mathcal{R}(x)$. The isolines depicted in Figure 1(c) correspond to the relationship $|\mathcal{O}_\mathcal{R}| = |\mathcal{O}^+| + |\mathcal{O}^-|$, which is true in Łukasiewicz logic (thus in particular also in Boolean logic).

6 Experimentally testing the approach

The proposed approach has been so far experimentally tested for six rules extraction methods on three benchmark data sets, as well as on data from one real-world knowledge discovery task [20]. For each method, 1–3 parameters were tuned, the values of them being chosen among 2–10 possibilities. For some data sets, some combinations of parameter values did not extract any rules. Whenever a particular combination of parameter vaules extracted a nonempty ruleset from the considered data, it was tested on those data by means of a 10-fold crossvalidation. Consequently, the number of rulesets extracted from each data set varied between 1000 and 1500.

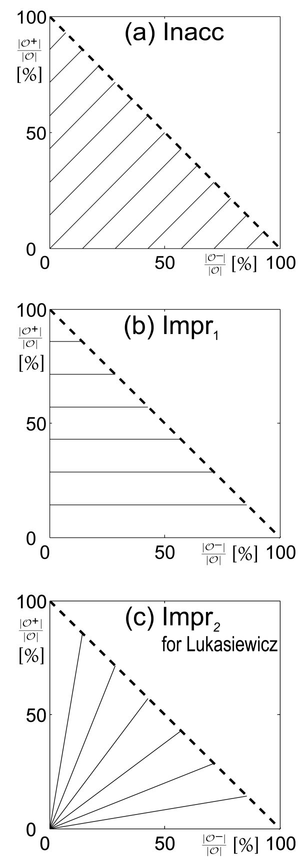

Fig. 1. Isolines of the three measures introduced in (20)–(22), drawn with respect to the coordinates $(\frac{|\mathcal{O}^-|}{|\mathcal{O}|}, \frac{|\mathcal{O}^+|}{|\mathcal{O}|})$ of points forming the proposed generalization of ROC curves
.

As a very brief illustration, Figure 2 shows the proposed generalization of ROC curves for two rulesets extracted from the best known benchmark set, the iris

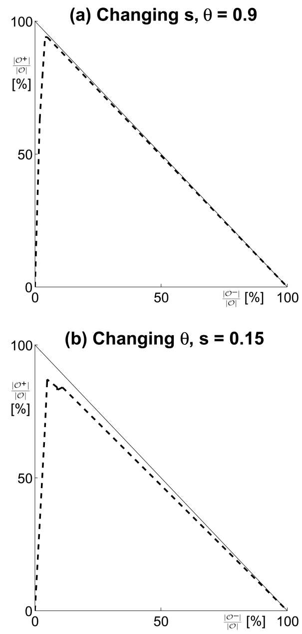

Fig. 2. Example of generalized ROC curves for rulesets extracted from the iris data by means of the GUHA quantifier founded implication

data, originally used in 1930s by R.A. Fisher [8], by means of the GUHA quantifier *founded implication.* This quantifier, denoted $\rightarrow_{s,\theta}$, $s, \theta \in (0, 1\rangle$ has its truth function $\mathrm{Tf}_{\rightarrow_{s,\theta}}$ defined in such a way that the rule $A_r \rightarrow_{s,\theta} C_r$ is valid exactly for those data for which the conditional probability $p(C_r|A_r)$ of the validity of C_r conditioned on A_r, estimated with the unbiased estimate $\frac{a}{a+b}$, is at least θ, whereas A_r and C_r are simultaneously valid in at least the proportion s of the data [13]. Hence, $\mathrm{Tf}_{\rightarrow_{s,\theta}} = 1$ iff $\frac{a}{a+b} \geq \theta$ & $\frac{a}{a+b+c+d} \geq s$. As was pointed out in [14], rules with this quantifier are actually association rules with support s and confidence θ. Each curve corresponds to changing only one of the parameters s, θ, the value of the other is fixed.

7 Conclusions

The paper has dealt with quality measures of rules extracted from data, though not in the usual context of individual rules, but in the context of whole rulesets. Three kinds of extensions of measures already in use for classification rulesets have been proposed. In addition, the concept of ROC-curves has been generalized, to enable investigating the dependence of general rulesets on the values of parameters of the extraction method.

The paper actuallly discusses some general aspects related to an ongoing investigation into the possibility to reflect uncertain validity of rulesets extracted from data when measuring their quality. The outcomes of that investigation are intended to be published elsewhere [17]. They comprise theoretical elaboration of the last proposed kind of extensions of ruleset quality measures, as well as results of extensive experimental tests on rulesets extracted from benchmark and real-world data sets by means of six methods attempting to cover a possibly broad spectrum of rules extraction methods. Those results indicate that the approach is feasible and can contribute to the ultimate objective of quality measures: to allow comparing the knowledge extracted with different data mining methods and investigating how the extracted knowledge depends on the values of their parameters.

Acknowledgment

The research reported in this paper has been supported by the grant No. 201/08/1744 of the Grant Agency of the Czech Republic and partially supported by the Institutional Research Plan AV0Z10300504.

References

1. R. Agrawal, H. Mannila, R. Srikant, H. Toivonen, and A.I. Verkamo. Fast discovery of association rules. In *Advances in Knowledge Discovery and Data Mining*, pages 307–328. AAAI Press, Menlo Park, 1996.
2. S.D. Bay and M.J. Pazzani. Detecting group differences. mining contrast sets. *Data Mining and Knowledge Discovery*, 5:213–246, 2001.
3. L. Breiman, J.H. Friedman, R.A. Olshen, and C.J. Stone. *Classification and Regression Trees.* Wadsworth, Belmont, 1984.
4. P. Clark and R. Boswell. Rule induction with CN2: Some recent improvements. In *Machine Learning – EWSL-91*, pages 151–163. Springer Verlag, New York, 1991.
5. L. De Raedt. *Interactive Theory Revision: An Inductive Logic Programming Approach.* Academic Press, London, 1992.
6. D. Dubois, Hüllermeier, and H. Prade. A systematic approach to the assessment of fuzzy association rules. *Data Mining and Knowledge Discovery*, 13:167–192, 2006.
7. W. Duch, R. Adamczak, and K. Grabczewski. A new methodology of extraction, optimization and application of crisp and fuzzy logical rules. *IEEE Transactions on Neural Networks*, 11:277–306, 2000.
8. R.A. Fisher. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 7:179–188, 1936.
9. A.A. Freitas. *Data Mining and Knowledge Discovery with Evolutionary Algorithms.* Springer Verlag, Berlin, 2002.

10. J. Fürnkranz and P.A. Flach. ROC 'n' rule learning – towards a better understanding of covering algorithms. *Machine Learning*, 58:39–77, 2005.
11. L. Geng and H.J. Hamilton. Choosing the right lens: Finding what is interesting in data mining. In F. Guillet and H.J. Hamilton, editors, *Quality Measures in Data Mining*, pages 3–24. Springer Verlag, Berlin, 2007.
12. P. Hájek. *Metamathematics of Fuzzy Logic*. Kluwer Academic Publishers, Dordrecht, 1998.
13. P. Hájek and T. Havránek. *Mechanizing Hypothesis Formation*. Springer Verlag, Berlin, 1978.
14. P. Hájek and M. Holeňa. Formal logics of discovery and hypothesis formation by machine. *Theoretical Computer Science*, 292:345–357, 2003.
15. D.J. Hand. *Construction and Assessment of Classification Rules*. John Wiley and Sons, New York, 1997.
16. R.J. Hilderman and T. Peckham. Statistical methodologies for mining potentially interesting contrast sets. In F. Guillet and H.J. Hamilton, editors, *Quality Measures in Data Mining*, pages 153–177. Springer Verlag, Berlin, 2007.
17. M. Holeňa. Measures of ruleset quality capable to represent uncertain validity. Submitted to *International Journal of Approximate Reasoning*.
18. M. Holeňa. Fuzzy hypotheses for Guha implications. *Fuzzy Sets and Systems*, 98:101–125, 1998.
19. M. Holeňa. Fuzzy hypotheses testing in the framework of fuzzy logic. *Fuzzy Sets and Systems*, 145:229–252, 2004.
20. M. Holeňa. Neural networks for extraction of fuzzy logic rules with application to EEG data. In B. Ribeiro, R.F. Albrecht, and A. Dobnikar, editors, *Adaptive and Natural Computing Algorithms*, pages 369–372. Springer Verlag, Wien, 2005.
21. M. Holeňa. Piecewise-linear neural networks and their relationship to rule extraction from data. *Neural Computation*, 18:2813–2853, 2006.
22. J.S.R. Jang. ANFIS: Adaptive-network-based fuzzy inference system. *IEEE Transactions on Systems, Man, and Cybernetics*, 23:665–685, 1993.
23. J.S.R. Jang and C.T. Sun. Neuro-fuzzy modeling and control. *The Proceedings of the IEEE*, 83:378–406, 1995.
24. K.A. Kaufman and R.S. Michalski. An adjustable description quality measure for pattern discovery using the AQ methodology. *Journal of Intelligent Information Systems*, 14:199–216, 2000.
25. E.P. Klement, R. Mesiar, and E. Pap. *Triangular Norms*. Kluwer Academic Publishers, Dordrecht, 2000.
26. S. Lallich, O. Teytaud, and E. Prudhomme. Association rule interestingness: Measure and statistical validation. In F. Guillet and H.J. Hamilton, editors, *Quality Measures in Data Mining*, pages 251–275. Springer Verlag, Berlin, 2007.
27. P. Lenca, B. Vaiilant, P. Meyer, and S. Lalich. Association rule interestingness meaures: Experimental and theoretical studies. In F. Guillet and H.J. Hamilton, editors, *Quality Measures in Data Mining*, pages 51–76. Springer Verlag, Berlin, 2007.
28. L. Lerman and J. Azè. Une mesure probabiliste contextuelle discriminante de qualite des règles d'association. In *EGC 2003: Extraction et Gestion des Connaissances*, pages 247–263. Hermes Science Publications, Lavoisier, 2003.
29. K. McGarry. A survey of interestingness measures for knowledge discovery. *Knowledge Engineering Review*, 20:39–61, 2005.
30. R.S. Michalski. Knowledge acquisition through conceptual clustering: A theoretical framework and algorithm for partitioning data into conjunctive concepts. *International Journal of Policy Analysis and Information Systems*, 4:219–243, 1980.
31. R.S. Michalski and K.A. Kaufman. Learning patterns in noisy data. In *Machine Learning and Its Applications*, pages 22–38. Springer Verlag, New York, 2001.
32. S. Mitra and Y. Hayashi. Neuro-fuzzy rule generation: Survey in soft computing framework. *IEEE Transactions on Neural Networks*, 11:748–768, 2000.
33. S. Muggleton. *Inductive Logic Programming*. Academic Press, London, 1992.
34. D. Nauck. Fuzzy data analysis with NEFCLASS. *International Journal of Approximate Reasoning*, 32:103–130, 2002.
35. D. Nauck and R. Kruse. NEFCLASS-X: A neuro-fuzzy tool to build readable fuzzy classifiers. *BT Technology Journal*, 3:180–192, 1998.
36. V. Novák, I. Perfilieva, A. Dvořák, C.Q. Chen, Q. Wei, and P. Yan. Mining pure linguistic associations from numerical data. To appear in International Journal of Approximate Reasoning.
37. J. Quinlan. *C4.5: Programs for Machine Learning*. Morgan Kaufmann Publishers, San Francisco, 1992.
38. A.B. Tickle, R. Andrews, M. Golea, and J. Diederich. The truth will come to light: Directions and challenges in extracting rules from trained artificial neural networks. *IEEE Transactions on Neural Networks*, 9:1057–1068, 1998.
39. H. Tsukimoto. Extracting rules from trained neural networks. *IEEE Transactions on Neural Networks*, 11:333–389, 2000.
40. M.J. Zaki, S. Parathasarathy, M. Ogihara, and W. Li. New parallel algorithms for fast discovery of association rules. *Data Mining and Knowledge Discovery*, 1:343–373, 1997.